Course Outline

Course Code	RSM 317 H1 S
Course Name	Text Mining and Natural Language Processing
Term, Year	Winter, 2025
Course Meets	Thu, 09:00 – 11:00
Web page URL	https://q.utoronto.ca Make sure you **always read the online Announcements**!

Instructor Details

Name	Email	Phone	Office Hours	Virtual Office Link
Gerhard Trippen ----------- TA:	Gerhard.Trippen@rotman.utoronto.ca (Please start subject with **RSM317)**	416-978-0383	Thu, 11:00 – 12:00 by appointment only and other times by appointment.	in-person mechanism preferred
TBD	TBD	NA	NA	NA

Course Description

This course will introduce the students to a diverse collection of text mining techniques and natural language processing using machine learning. These techniques are often aimed at identifying and quantifying various structures in the text data to answer business problems and provide managerial insights. Model validation and effective communication of model-based results will be stressed. The course will employ a "white-box" methodology, which emphasizes an understanding of the algorithmic and statistical model structures and how they apply to text analysis. This is a hands-on course focused on developing practical skills across a broad range of the most common text mining and natural language processing techniques. Following leading industry standards, the course will use Python to apply a number of different algorithms to real-world big text data.

Learning Outcomes

"Learn how to process, classify, cluster, summarize, understand syntax, semantics and sentiment of text data with the power of Python!"

By the end of this course, students will be able to:
- Process and analyze text data: Demonstrate proficiency in applying text preprocessing techniques, including tokenization, normalization, and spellchecking, to prepare textual data for analysis.
- Identify and quantify structures in text data: Apply methods such as Bag of Words, TF-IDF, and document-term matrix representations to extract meaningful patterns and structures from textual content.

- Develop classification models: Build and evaluate text classification models using machine learning techniques, including neural networks, for tasks such as sentiment analysis and categorization.
- Perform clustering and topic modeling: Apply clustering algorithms and topic modeling techniques to discover latent patterns, such as document groupings and topic structures, within large text datasets.
- Summarize and extract information: Use text summarization techniques and information extraction methods to distill insights from large corpora of textual data.
- Analyze semantic and syntactic text characteristics: Assess semantic meaning, syntax, and sentiment in textual data to derive actionable business insights.
- Generate and validate models: Apply model validation techniques to ensure the robustness of text-based machine learning models and communicate results effectively to stakeholders.
- Leverage Python for text analysis: Utilize Python libraries (e.g., NLTK, spaCy, scikit-learn) to implement text mining and NLP techniques on real-world datasets.
- Apply NLP solutions to business problems: Demonstrate the ability to translate unstructured text data into managerial insights to address specific business challenges.
- Engage with Generative AI (GenAI): Explore and implement text generation methods using modern generative AI approaches.

Course Prerequisites
CSC108H1

Course Materials

Required Readings

Reading Package	DataCamp	Required	https://www.datacamp.com/
Textbook	Text Analytics with Python: A Practitioner's Guide to Natural Language Processing Paperback, 2019, by Dipanjan Sarkar	Optional	
Textbook	Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data, 2016, by Dipanjan Sarkar	Optional	
Additional Reading	Applied Text Analysis with Python Enabling Language-Aware Data Products with Machine Learning By: Benjamin Bengfort; Rebecca Bilbro; Tony Ojeda Publisher: O'Reilly Media Print ISBN: 9781491963043, 1491963042Print ISBN: 9781491963043, 1491963042 eText ISBN: 9781491962992, 1491962992	Optional	
Additional Reading	Python for Data Analysis, 3E by Wes McKinney (pub. yr. 2022) Free online access: https://wesmckinney.com/book/	Optional	Online

Electronic Course Materials
This course will be using the following electronic course materials:

DataCamp: https://www.datacamp.com/

Anaconda: https://www.anaconda.com/

These materials will cost a total of **$0.00**. The use of these materials complies with all University of Toronto policies which govern fees for course materials.

Evaluation and Grades
Grades are a measure of the knowledge and skills developed by a student within individual courses. Each student will receive a grade on the basis of how well they have command of the course materials, skills and learning objectives of the course.

Work	Percentage of grade	Due Date
Individual Coding Assignments	10%	Ongoing
Group Assignment 1	15%	2025-02-13 23:59
Group Assignment 2	15%	2025-03-13 23:59
Group Assignment 3	15%	2025-04-03 23:59
Final Exam	45%	TBD

Course Format and Expectations
We will be using Jupyter Notebook / DataCamp for data analysis exercises throughout the course. All students will receive a **free** DataCamp subscription for six months, with access to all the online courses DataCamp offers (not just the chapters that are covered in the course).
To register for DataCamp an email address will be required which you need to share with me in the beginning of the term. For students who would prefer to opt out and not use DataCamp, solutions to exercises can be submitted to me by email.

Writing Assignments or Presentations
The case studies are intended to help you intended to help you develop your communication skills. How well you communicate your ideas, in writing or orally, will be considered in the evaluation of the assignment. In your written assignments, you should aim for clarity, strong organization, concision, professionalism, and correct grammar. Your presentations should reflect strong planning and organization, clarity of speech, and an engaging demeanour. Sources, whether in written or presentation assignments, should always be correctly attributed.

Support is available through the RC Centre for Professional Skills (CPS) for students who would like help or feedback on their writing or speaking (presentations). CPS offers both individual and group appointments with trained writing instructors and presentation coaches who are familiar with the RC program and common types of business assignments. You can also access your college Writing Centres for help with written assignments.

You can book an appointment with a writing or presentation coach through the RC Centre for Professional Skills Writing Centre. For more information about writing centres, student supports, and study resources, see the Writing and Presentation Coaching academic support page.

Team or Group Assignments
The case studies require students to work in teams of 3-4 (preferably 4). Learning to work together in teams is a crucial transferrable skill you will use not only in your coursework, but also in your future careers. Support is available if you encounter common teamwork challenges such as:

- Team members feeling left out of the team.
- Team members not responding in a timely manner to communication.
- Division or quality of work among team members being unequal or unfair.

Consult the [Centre for Professional Skills Teamwork Resources page](#) for tips, strategies, and best practices. You can also [book an appointment with a teamwork mentor](#) through the RC Centre for Professional Skills Writing Centre. Teamwork mentors can help you resolve or mitigate conflict, strategize on planning, or improve team communication.

If you are a student registered with Accessibility Services, and extensions are one of your academic accommodations, consult with your Accessibility Advisor about the teamwork in this course.

Missed Tests and Assignments

Students who miss a term test or assignment for reasons entirely beyond their control (e.g. illness) may request special consideration **within 2 business days** of the missed midterm/test/assignment due date.

In such cases, students must:

1. Complete the Request for Special Consideration form: https://uoft.me/RSMConsideration
2. Provide documentation to support the request, eg. Absence Declaration from [ACORN](#), medical note etc.

Please note: As of September 2023, students may use the Absence Declaration on ACORN *one time per term* to report an absence and request consideration. **Any subsequent absence will require a [Verification of Illness form](#) or other similar relevant documentation.**

Students who do not submit their requests and documentation within 2 days may receive a grade of 0 (zero) on the missed course deliverable.

Final Exams: If you miss the final exam in this course for a legitimate reason (illness, etc) you will need to contact your College Registrar to file a petition for a deferred exam. This deferred exam will be written at a later date as established by the Faculty of Arts & Science. Instructions can be found here: https://www.artsci.utoronto.ca/current/faculty-registrar/petitions-appeals/preparing-petition

Late Assignments

All assignments are due on the date and at the time specified in Quercus. Late submissions will normally be penalized by 20% if the assignment is not received on the specified date, at the specified time. A further penalty of 10% will be applied to each subsequent day.

Statement on Equity, Diversity and Inclusion

The University of Toronto is committed to equity, human rights and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect where all members of our community can express themselves, engage with each other, and respect one another's differences. U of T does not condone discrimination or harassment against any persons or communities.

Commitment to Accessibility

The University is committed to inclusivity and accessibility, and strives to provide support for, and facilitate the accommodation of, individuals with disabilities so that all may share the same level of access to opportunities and activities offered at the University.

If you require accommodations for a temporary or ongoing disability or health concern, or have any accessibility concerns about the course, the classroom or course materials, please email Accessibility Services or visit the Accessibility Services website for more information as soon as possible. Obtaining your accommodation letter may take up to several weeks, so get in touch with them as soon as possible. If you have general questions or concerns about the accessibility of this course, you are encouraged to reach out to your instructor, course coordinator, or Accessibility Services.

Plagiarism Detection

Normally, students will be required to submit their course essays to the University's plagiarism detection tool for a review of textual similarity and detection of possible plagiarism. In doing so, students will allow their essays to be included as source documents in the tool's reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of this tool are described on the University's Plagiarism Detection Tool FAQ page from Centre for Teaching Support & Innovation.

Moss - Measure Of Software Similarity:

Moss is an automatic system for determining the similarity of programs. Moss will be used for reviewing source code and preventing possible plagiarism.

Generative AI / ChatGPT

The knowing use of generative artificial intelligence tools, including ChatGPT and other AI writing and coding assistants, for the completion of, or to support the completion of, an examination, term test, assignment, or any other form of academic assessment, may be considered an academic offense in this course.

Academic Integrity

Academic Integrity is a fundamental value essential to the pursuit of learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the U of T degree that you earn will continue to be valued and respected as a true signifier of a student's individual work and academic achievement. As a result, the University treats cases of academic misconduct very seriously.

The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic misconduct, the process for addressing academic offences and the penalties that may be imposed. You are expected to be familiar with the contents of this document. Potential offences include, but are not limited to:

In papers and assignments

- Using someone else's ideas or words without appropriate acknowledgement.
- Submitting your own work in more than one course without the permission of the instructor.
- Making up sources or facts.
- Obtaining or providing unauthorized assistance on any assignment (this includes collaborating with others on assignments that are supposed to be completed individually).

On test and exams

- Using or possessing any unauthorized aid, including a cell phone.

- Looking at someone else's answers.

- Misrepresenting your identity.

- Submitting an altered test for re-grading.

Misrepresentation

- Falsifying institutional documents or grades.

- Falsifying or altering any documentation required by the University, including (but not limited to) medical notes.

All suspected cases of academic dishonesty will be investigated by the procedures outlined in the *Code of Behaviour on Academic Matters*. If you have any questions about what is or is not permitted in the course, please do not hesitate to contact the course instructor. If you have any questions about appropriate research and citation methods, you are expected to seek out additional information from the instructor or other U of T or RC resources such as the RC Centre for Professional Skills, the College Writing Centres or the Academic Success Centre.

Email

At times, the course instructor may decide to communicate important course information by email. As such, all U of T students are required to have a valid UTmail+ email address. You are responsible for ensuring that your UTmail+ email address is set up and properly entered on ACORN. For more information visit the Information Commons Help Desk.

Forwarding your utoronto.ca email to a Gmail or other type of email account is not advisable. In some cases, messages from utoronto.ca addresses sent to Gmail accounts are filtered as junk mail, which means that important messages from your course instructor may end up in your spam or junk mail folder.

Recording Lectures

Lectures and course materials prepared by the instructor are considered by the University to be an instructor's intellectual property covered by the Canadian Copyright Act. Students wishing to record a lecture or other course material in any way are required to ask the instructor's explicit permission and may not do so unless permission is granted. Students who have been previously granted permission to record lectures as an accommodation for a disability are excepted. This includes tape recording, filming, photographing PowerPoint slides, Quercus materials, etc.

If permission for recording is granted by the instructor (or via Accessibility Services), it is intended for the individual student's own study purposes and does not include permission to "publish" them in any way. It is forbidden for a student to publish an instructor's notes to a website or sell them in any other form without formal permission.

Weekly Schedule *(TENTATIVE)*

Session	Date	Topic	Readings (Dipanjan Sarkar, 2019)
1	2025-01-09	**Introduction and Python Refresher** Overview and Working with Text Data	Chapter 1: *Introduction to Text Analytics and NLP* Chapter 2: *Python Basics*
2	2025-01-16	**Natural Language Processing Basics**	Chapter 3: *Processing and Understanding Text*
3	2025-01-23	**Regular Expressions and Text Preprocessing**	Chapter 4: *Text Cleaning and Preprocessing*
4	2025-01-30	**Tokenization and Normalization**	Chapter 5: *Tokenization, Stemming, and Lemmatization*
5	2025-02-06	**Feature Engineering: Bag of Words and TF-IDF**	Chapter 6: *Feature Extraction from Text*
6	2025-02-13	**Text Classification with Neural Networks**	Chapter 7: *Text Classification and Supervised Learning*
7	2025-02-27	**Word Embeddings: Word2Vec**	Chapter 8: *Word Embeddings and Feature Representation*
8	2025-03-06	**Text Summarization and Topic Modeling**	Chapter 9: *Topic Modeling and Text Summarization*
9	2025-03-13	**Text Similarity and Clustering**	Chapter 9: *Topic Modeling and Text Summarization* (Clustering section)
10	2025-03-20	**Semantic Analysis**	Chapter 10: *Semantic Analysis and Understanding Context*
11	2025-03-27	**Sentiment Analysis with Deep Learning**	Chapter 11: *Sentiment Analysis with Machine Learning*
12	2025-04-03	**Text Generation (GenAI) and Transformers**	Chapter 12: *Text Generation and Sequence Models*

Please note that the last day you can drop this course without academic penalty is March 10, 2025.

Other Useful Links

- [Become a volunteer note taker](#)
- [Accessibility Services Note Taking Support](#)
- [Credit / No-Credit in RSM courses](#)
- [Rotman Commerce Academic Support](#)
- [Where to find teaching assistant opportunities](#)

URL links for print

- ACORN: http://www.acorn.utoronto.ca/
- Email Accessibility Services: accessibility.services@utoronto.ca
- Accessibility Services website: http://studentlife.utoronto.ca/as
- University's Plagiarism Detection Tool FAQ: https://uoft.me/pdt-faq
- The University of Toronto's Code of Behaviour on Academic Matters: http://www.governingcouncil.utoronto.ca/policies/behaveac.htm
- Information Commons Help Desk: http://help.ic.utoronto.ca/category/3/utmail.html
- Become a volunteer note taker: https://studentlife.utoronto.ca/program/volunteer-note-taking/
- Accessibility Services Note Taking Support: https://studentlife.utoronto.ca/service/note-taking-support/
- Credit / No-Credit in RSM courses: https://rotmancommerce.utoronto.ca/current-students/degree-requirements/credit-no-credit-option/
- Rotman Commerce Academic Support: https://rotmancommerce.utoronto.ca/current-students/academic-support/
- Book an appointment with a writing or presentation coach: http://uoft.me/writingcentres
- Writing and Presentation Coaching academic support page: https://rotmancommerce.utoronto.ca/current-students/academic-support/writing-and-presentation-coaching/
- Centre for Professional Skills Teamwork Resources page: https://rotmancommerce.utoronto.ca/teamwork-resources
- Book an appointment with a Teamwork Mentor: http://uoft.me/writingcentres